The Wilson Effect: The Increase in Heritability of IQ With Age

Thomas J. Bouchard Jr.
Psychology Department, University of Minnesota, Minneapolis, USA

Ronald Wilson presented the first clear and compelling evidence that the heritability of IQ increases with age. We propose to call the phenomenon ‘The Wilson Effect’ and we document the effect diagrammatically with key twin and adoption studies, including twins reared apart, that have been carried out at various ages and in a large number of different settings. The results show that the heritability of IQ reaches an asymptote at about 0.80 at 18–20 years of age and continuing at that level well into adulthood. In the aggregate, the studies also confirm that shared environmental influence decreases across age, approximating about 0.10 at 18–20 years of age and continuing at that level into adulthood. These conclusions apply to the Westernized industrial democracies in which most of the studies have been carried out.

Keywords IQ, age, intelligence, heritability, twins, adoptees

Monozygotic twins became increasingly concordant with age and also paralleled each other for the spurts and lags in development. Dizygotic twins became less concordant with age and eventually matched their singleton siblings as closely as one another. The overall results suggested that the course of mental development is guided by the intrinsic scheduling of the genetic program acting in concert with maturational status and environmental influence. (Wilson, 1978, p. 439).

The fact that the heritability of IQ increases with age is now a widely accepted phenomenon (Boomsma et al., 2002; Deary, 2012; Johnson, 2010b). Such acceptance, even recognition, was not always the case. Early studies focused mostly on children and seldom explored the effect of age. The classic review of family studies by Erlenmeyer-Kimling and Jarvik (1963) does not mention age effects on kin correlations nor (embarrassingly for this author) does the 1981 review by Bouchard and McGue (1981) or the review by Plomin and DeFries (1980). As late as 1997, Devlin et al. (1997) asserted there were no age effects on the heritability of IQ.

The purpose of this paper is to provide a striking illustration of this phenomenon and provide it with a name — ‘The Wilson Effect’ — in honor of Ronald Wilson, the developmental behavior geneticist who most clearly articulated its existence and emphasized the importance of putting the findings in both a developmental and evolutionary context (Wilson, 1978, 1983). For those not familiar with Wilson’s work, Plomin (1987) provides a thoughtful summary of his career.

Figure 1 from Wilson (1983) is a plot of kin correlations across age. The data were gathered by the justly famous Louisville Longitudinal Twin Study. The plot illustrates the changes in kin correlations that mediate the effect; relatively stable monozygotic twin (MZ) correlations from 2 years of age and decreasing same-sex dizygotic twin (DZ) correlations after 4 years of age. The converging of sibling-twin correlations with the DZ correlations at 6 years of age and the increasing age-to-age correlations complement the twin correlations. The effect is well defined by 5–6 years of age when the age-to-age correlations approximate the MZ correlations.

Figure 2 puts together a number of key studies (large and representative twin samples and some smaller but statistically powerful adoption designs) that illustrate the generality of the Wilson Effect.

The Dutch data are from Boomsma et al. (2002) and consist of eight cross-sectional twin studies that provide a frame of reference for the other studies. For illustrative purposes, a polynomial curve (solid black line) has been fit to the estimates of genetic influence (heritabilities) and it reaches an asymptote at about age 20. The estimates of shared environmental influence (common environment)
FIGURE 1
Mental development correlations for MZ twins, DZ twins, twin–sibling sets, parent–offspring sets; and for each child with itself, age to age.

FIGURE 2
Estimates of genetic and shared environmental influence on g by age. The age scale is not linear (see text for details).
show a linear decrease (black squares) from about 0.55 at age 5 to 0.00 at age 12 and remain at 0.00 through 50 years of age. These findings are from an advanced industrialized society with a sophisticated social welfare system and provide a baseline against which other data can be compared. All genetic estimators in this study and those to be added below are of the broad heritability \( (h^2) \), which includes additive genetic variance, dominance genetic variance, and genetic variance due to assortative mating (Vinkhuyzen et al., 2012; Visscher et al., 2008). Common or shared environmental estimates \( (c^2) \) differ slightly depending on the particular design, but the values are all quite low so it would take very large samples to partition any differences. Most twin studies set \( c^2 \) to zero if the estimate is not statistically significant. This is an unfortunate practice in our opinion.

How do the Dutch results compare to data sets generated in other settings? Haworth et al. (2010) combined the raw IQ data from six large twin studies (three from the United States, one from the United Kingdom, one from Australia, and one from the Netherlands). These data were divided into three age groups; mean ages of 9, 12, and 17 years. The findings are shown as numbers with no associated coding in the figure \( (h^2 = 0.41, 0.55, \text{ and } 0.66; c^2 = 0.33, 0.28, \text{ and } 0.16) \). The heritabilities go up with age, and the line runs parallel with the Dutch data but lies about 10% below it. There may be some overlap in the Dutch sample in this collection of samples with the Boomsma et al. (2002) Dutch data.

What happens if we add data from monozygotic twins reared apart (MZA)? These twin correlations estimate the broad heritability directly (Bouchard, 2009). The mean of five studies of adults is 0.74. The line from the Haworth data at age 17 is projected \( (\ldots \rightarrow) \) to the MZA data and it largely parallels the Dutch data. We can also estimate shared environmental influence on adult IQ by subtracting the adult MZA correlation from the monozygotic twins reared together (MZT) correlation. An estimate of the MZT correlation in advanced industrialized countries (Norway, 1181 pairs, age 18 to 21) is 0.81 (Sundet et al., 2005), although an earlier sample suggests 0.84 (Sundet et al., 1988). Devlin et al. (1997) updated the Bouchard and McGue (1981) estimate of 0.86 and reported a correlation of 0.85. Finkel et al. (1995) report 0.80 and 0.84 for middle-aged Minnesota and Swedish samples. We use 0.84. This gives an estimate of 0.10 for shared environment and that value has been entered into Figure 2. The results for shared environment are almost a mirror image of the genetic results.

There are a variety of arguments commonly made against twin studies (Bouchard & McGue, 2003). One is that they are volunteer samples and do not represent the entire range of environments experienced by the population to which we wish to generalize. Ian Deary’s group in Edinburgh, Scotland has addressed this question. They have shown ‘Large, consistent estimates of the heritability of cognitive ability in two entire populations of 11-year-old twins from Scottish Mental Survey of 1932 and 1947’ (Benjamin et al., 2005). The estimates are \( h^2 = 0.70 \text{ and } c^2 = 0.21 \). The large and intentionally population-representative Minnesota Twin Family Study carried out a cognitive ability assessment at age 11 as well. The results were almost identical, \( h^2 = 0.67 \text{ and } c^2 = 0.26 \) (Johnson et al., 2005). These combined data are shown as filled boxes on the line drawn at age 11 years in Figure 2. There is also a Norwegian data set based on military recruits (again a comprehensive sample that includes all males tested in their late teens to early twenties) gathered from 1931 to 1960 (Sundet et al., 1988). When we add those values, 0.68 and 0.16 and shown in braces \( [] \), they fall right on the lines connecting the Haworth and MZA data.

The very modest shared environmental influence by mid-childhood revealed by twin studies always puzzles psychologists, as they intuitively expect siblings to become more similar as they spend more time together in the same family and its social and educational environment. Therefore, it is important to subject this counter-intuitive finding to independent tests. The most powerful design for estimating this effect is to study unrelated siblings raised in the same family (URT). When we examine this literature, we find an estimate of 0.26 in childhood (13 studies) and an estimate of 0.04 in adulthood (Bouchard, 2009, Figure 5). We note here that Devlin et al. (1997) found these values to be extremely heterogeneous and dropped them from their analysis. There are only five studies in adulthood; estimates from four cluster are very close to zero and one (a United States study with participants at about age 20 is 0.19). Three of the five studies are longitudinal and all show the decline characteristic of the twin studies in the figure. The childhood and adulthood data points are shown in Figure 2 as two asterisks connected with an arrow pointed to the later age. There are no twins in these studies. Thus, different methods (constructive replication) lead to the same conclusion and this conclusion does not appear to be driven by the assumptions underlying twin studies. A second way to estimate shared environmental influence is by using a comprehensive adoption sample, not just siblings. The Texas Adoption Project (TAP; Horn & Loehlin, 2010) used 14 kinships drawn from about 300 families and estimated a shared environmental influence of 0.00 at age 17. As with the URT studies, there are no twins in this project. TAP also provides us with an estimate of heritability. The TAP values are shown in brackets \( [] \), 0.78 for heritability and 0.00 for shared environmental influences; these values are very consistent with the Dutch data at the same age.

There is another adoption study that allows us to estimate both heritability and shared environmental influence. It is a small sample study, but because it used Danish military data it is highly representative of the Danish male population (Teasdale & Owen, 1984). The study made use of siblings reared apart, maternal and paternal half-siblings reared apart, unrelated reared-together, and full siblings
reared together. Four of these five kinships constitute the entire population of such individuals rather than samples; the last group (full siblings reared together) is representative but not a population sample. The estimates are shown in a chart in Figure 2 and the values are 0.96 and 0.02. This study provides the highest heritability estimate ever reported and nearly the lowest estimate of shared environmental influence. It did not make use of twins. The unrelated reared together from this study are included in the adult URT sample shown with an asterisk.

Finally, we have estimates of heritability and shared environment from a sample of 65-year-old MZ and DZ twins reared apart and together from Sweden (Reynolds et al., 2005). The estimates are 0.91 and 0.00 and shown outside the chart. They are highly consistent with the Dutch data. We show the Reynolds et al. data because these come from the only study of older twins that included both twins reared together and twins reared apart. Lee et al. (2010) have reviewed genetic influence on cognitive abilities in the elderly (over age 65). They concluded that ‘cross-sectional studies seem to suggest a decrease in heritability of g, from 80% at the mean age of 65 to 60% at the mean age of 82’ (p. 10). Figure 2 is consistent with a great deal of other data not included. Examples would be:

1. The Louisville Twin Study, a longitudinal study of twins reared together (Matheny, 1990; Wilson, 1983).
2. Other twin studies including one in Russia (Malykh et al., 2005). For a recent review of twin studies, see Segal and Johnson (2009) and for recent studies, see (Bratko et al., 2010; Davis et al., 2009).
3. Very large studies that include both twin and singleton data (Calvin et al., 2012).
4. Virtual twin (unrelated reared-together of highly similar childhood ages) studies, both cross-sectionally and longitudinally (Segal et al., 2012).
5. The Colorado Adoption Project (CAP), a longitudinal study of adoptive and biological families (Plomin et al., 1997).
6. Other contemporary adoption studies (Petrill & Deater-Deckard, 2004; Plug & Vijverberg, 2003; Scarr, 1997; Scarr & Weinberg, 1978) as well as much older ones (Burks, 1928, 1938; Leahy, 1935). For a detailed discussion of these and other earlier studies, see Bouchard (1993), Bouchard et al. (1996), and Bouchard and Segal (1985).

The observed changes cannot be explained by changes in brain size, as brain volume does not increase over the age period from 8 to 22 years of age (De Bellis et al., 2001; Giedd et al., 1999), although the relative sizes of different parts do change. Cerebral hemisphere size decreases from age 20 to 80 (Raz et al., 2004). Regarding changes in the brain with age and their association with IQ, see also (Karama et al., 2011; Lenroot et al., 2009). Brain volume is highly heritable even in older twins (Lessov-Schlaggar et al., 2012).

What does Figure 2 tell us? It tells us three big things, all of which were intuitively plausible to many psychologists and consequently adamantly denied for many years (Gould, 1981, 1996; Kamin, 1974; Lewontin et al., 1984).

The first is that IQ is significantly heritable beginning at least at age 7. The second is that by age 10 genetic variance is larger than shared environmental variance. Third, the heritability increases with age until late adulthood when it reaches the mid-seventies and higher for some populations. Contrary to the widespread belief that ‘the slings and arrows of outrageous fortune’ accumulate over time (a version of the spun glass theory of mind; Schofield, 1986), the brain appears to be a robust and resilient mechanism.

The picture does change very late in life, but as one would expect attrition changes the characteristics of the sample — sample selection bias (Johnson et al., in press; Rabbit & Lunn, 2008). Nevertheless, heritability is still quite high (0.62) for 80-year-old twins when both individuals are free of major motor, sensory, or cognitive impairment (McClearn et al., 1997).

Currently, much emphasis is being placed on SES by heritability interactions (Nisbett et al., 2012). The results of the various studies are mixed, with some failure to replicate and some results in the opposite direction. It may be that the effect is real in childhood and young adulthood but fades in adulthood. As Nisbett et al. pointed out: ‘studies of adults must contend with the low magnitude of shared environmental components overall, and it may be difficult to detect interactions when there are low shared environment effects to work with’ (p. 3).

It may also be that different genes are at work in adulthood and the processes they mediate do not interact with SES (or whatever SES indicates). That is, the SES × heritability interaction may be transient. Wilson’s (1978) work demonstrating spurt and lags in mental development clearly suggests that the development of IQ is strongly influenced by an intrinsic genetic program that expresses itself differently at different ages. In addition, Moffitt et al. (1993), using a large longitudinal and representative sample (the ‘Dunedin Sample’ measured at ages, 7, 9, 11, and 13), have shown that ‘in a non-trivial minority of children, naturalistic IQ change is marked and real, but this change is variable in its timing, idiosyncratic in its source, and transient in its course’ (p. 455). They argue that while IQ can be thrown off its trajectory, IQ change ‘appears to conform to recovery curves and seems to reflect level-mainlining or even level-seeking phenomena. Intellectual performance that was reliably deflected across time was characterized by a self-righting tendency. IQ appeared to be elastic, rather than plastic’ (p. 496). ‘To our mind, the most compelling finding in the Moffitt et al. study was something they did not find, namely a pattern of change that consisted of a constantly declining IQ. Such a slope would be predicted from the ‘spun glass theory of mind’ (Bouchard, 1995). Moffitt et al. cite the work of Scarr and McCartney (1983).
regarding how genotypes (embedded in individuals) create their own environments as the best theory to explain their results. Figure 2 demonstrates that shared environmental influence is also transient to a considerable extent. Shared environmental influence declines monotonically with age, dropping from about 0.55 at 5 years of age (Dutch data) to 0.10 in adulthood. It is noteworthy that all other estimates after 10 years of age are higher than the zero estimates from the Dutch data. These findings, of course, hold for what might be called the normal range of environments as exemplified by the ‘Dunedin Sample’.

The findings supporting the ‘Wilson Effect’ are highly robust:

1. They do not depend on a single design. Adoption studies with various combinations of kinships, not including twins, yield the same results as twin studies.
2. The findings are much the same across a variety of Western industrialized countries.
3. Many of the samples are quite comprehensive so their results probably apply broadly.

The fundamental point, emphasized by Wilson, is that different genes are in play at different times and the manner in which they influence an organism’s transactions with the environment also varies with time. This possibility is now recognized (Haworth et al., 2010) but still not widely appreciated, though concrete evidence for it remains scanty. Increasing heritability over time is a quite general phenomenon, having been observed in many phenotypes in both human and non-human species. J. P. Scott (1990), who carried out classic work on the behavior genetics of behavior in dogs with John Fuller (Scott & Fuller, 1965), reported that:

We thought that the best time to study the effects of genetics would be soon after birth, when behavior still had little opportunity to be altered by experience. On the contrary, we found that the different dog breeds were most alike as newborns; that is genetic variation in behavior develops postnatally, in part as a result of the timing of gene action and in part from the interaction of gene action and experience, social, and otherwise. (p. vii)

More generally, a number of human behavioral phenotypes show age-related changes in heritability (Bergen et al., 2007).

An important limitation of the presentation in Table 2 is that the meaning of the findings can be interpreted in a number of ways as the variance components have been standardized. A more complete picture would flow from the presentation of unstandardized variances. Such an analysis would, however, require a common, well-developed intelligence test be used at each age. Such instruments do exist and have been shown to have the same factor structure from age 6 to 70 (Bickley et al., 1995). A second limitation and source of variance in the findings is the great variety of instruments, their varying quality, and the number of tests used to estimate IQ. Nevertheless, most of the studies did use reasonably well-developed instruments with a variety of subtests and we know that such batteries do a pretty good job of estimating IQ if not the g factors itself (Johnson et al., 2008).

Ronald Wilson ended a number of his papers with a clear statement emphasizing that his findings did not imply that environments are unimportant. He used a variation of the following quote in three important papers (Wilson, 1977, 1978, 1983) and it is still well worth reproducing.

The contribution of the parents, whether natural or adoptive, is in potentiating the child’s inherent capabilities, in creating an atmosphere of enthusiasm for learning, and in adapting their expectations to the child’s capability. The wide diversity within families emphasizes the importance of giving each child full opportunity for development, and indeed of making sure that the opportunity is taken. The ultimate goal is the maximum realization of each child’s intelligence, coupled with a sense of satisfaction and personal accomplishment in its use. There is no better way to foster such development than by a supportive and appropriately stimulating family environment.

We conclude with our own less elegant commentary regarding how these findings should be understood and the role of the environment in the shaping of human psychological traits. It is important to specify the populations to which any results can be generalized and not misinterpret what they mean. The samples were drawn almost exclusively from Western industrial democracies. These settings have characteristic environments. Only a few of the participants were raised in real poverty or by illiterate parents, and all study participants had access to the contemporary educational programs typical of those societies. This is the domain to which we can generalize. The results do not mean that environments are irrelevant or unimportant. The proximate causes of variance in IQ are the ‘cognitively stimulating experiences’ that the individual is provided with and seeks for him/herself. Early in life, those experiences are primarily imposed (or not imposed) on the individual and that shows up as shared environmental influence. As the individual becomes more of an independent agent, the effective experiences are to a large extent self-selected. As Martin et al. (1986) put it: ‘humans are exploring organisms whose innate abilities and predispositions help them select what is relevant and adaptive from the range of opportunities and stimuli presented in the environment. The effects of mobility and learning, therefore, augment rather than eradicate the effects of the genotype on behavior’ (p. 4368). This idea has been elaborated in some detail by Scarr (1996) who has argued that people make their own environments1 and Bouchard (1996, 1997; Johnson, 2010a), who has elaborated the Experience Producing Drive Theory of Keith Hayes (1962).
Acknowledgments
The author thanks Wendy Johnson for constructive comments on a previous version of the paper.

Endnote
1 Scarr did not cite Hayes (1962) in any of her papers and I asked her about this. She expressed embarrassment and told me that indeed she had read the Hayes paper and regretted not having cited it.

References


